OIPE
MAR 1 8 2004
PATENT & TRADEMARK OFFICE

· 7
JM
4/ 14/04

BEST AVAILABLE COPY

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	09/788,269
Applicant	:	Jonathan W. Jarvik
Filed	:	February 16, 2001
Title	:	Methods and Products for Peptide-Based cDNA Characterization and Analysis
Group Art Unit	:	1645
Examiner	:	Not Yet Assigned

Mail Stop Petition
Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PETITION TO REVIVE UNDER 37 C.F.R. § 1.181

Sir:

Applicant has received the Notice of Abandonment dated January 21, 2004, in error. Although the Notice of Abandonment states that no response to the Notice to File Missing Parts mailed on April 6, 2001 was received, in fact a response to the Notice to File Missing Parts was timely filed on June 29, 2001. Copies of all the papers filed in response to the Notice to File Missing Parts, as filed on June 29, 2001, are attached, which copies include a copy of the postcard receipt dated July 3, 2001.

The Petitions Branch is asked to vacate the Notice of Abandonment dated January 21, 2004 and to return the above-identified application to the Examining Group for examination. The Petitions Branch is also requested to arrange if possible for the docketing of this application so that it can be taken up for examination without further delay, in view of the early filing date of the application.

<table>
<tr><td>

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to Mail Stop Petition, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on March 16, 2004.

_____Kimberly N. Welday_____
(Name of Person Mailing Paper)

Kimberly N. Welday   March 16, 2004
Signature   Date

</td><td>

Respectfully submitted,
WEBB ZIESENHEIM LOGSDON
ORKIN & HANSON, P.C.

By _____
Barbara E. Johnson
Registration No. 31,198
Attorney for Applicant
700 Koppers Building
436 Seventh Avenue
Pittsburgh, Pennsylvania 15219-1818
Telephone: 412-471-8815
Facsimile: 412-471-4094

</td></tr>
</table>

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
09/788,269	02/16/2001	Jonathan W. Jarvik	2087-010261

CONFIRMATION NO. 5283

ABANDONMENT/TERMINATION LETTER

Barbara E. Johnson
WEBB ZIESENHEIM LOGSDON ORKIN & HANSON
700 Koppers Building
436 Seventh Avenue
Pittsburgh, PA 15219-1818

OC000000011727589

Date Mailed: 01/21/2004

NOTICE OF ABANDONMENT UNDER 37 CFR 1.53 (f) OR (g)

The above-identified application is abandoned for failure to timely or properly reply to the Notice to File Missing Parts (Notice) mailed on 04/06/2001.

- No reply was received.

A petition to the Commissioner under 37 CFR 1.137 may be filed requesting that the application be revived.

Under 37 CFR 1.137(a), a petition requesting the application be revived on the grounds of **UNAVOIDABLE DELAY** must be filed promptly after the applicant becomes aware of the abandonment and such petition must be accompanied by: (1) an adequate showing of the cause of unavoidable delay; (2) the required reply to the above-identified Notice; (3) the petition fee set forth in 37 CFR 1.17(l); and (4) a terminal disclaimer if required by 37 CFR 1.137(d).

Under 37 CFR 1.137(b), a petition requesting the application be revived on the grounds of **UNINTENTIONAL DELAY** must be filed promptly after applicant becomes aware of the abandonment and such petition must be accompanied by: (1) a statement that the entire delay was unintentional; (2) the required reply to the above-identified Notice; (3) the petition fee set forth in 37 CFR 1.17(m); and (4) a terminal disclaimer if required by 37 CFR 1.137(d).

Any questions concerning petitions to revive should be directed to the "Office of Petitions" at (703) 305-9282. Petitions should be mailed to: Mail Stop Petitions, Commissioner for Patents, P.O. Box 1450, Alexandria VA 22313-1450.

*A copy of this notice **MUST** be returned with the reply.*

Customer Service Center
Initial Patent Examination Division (703) 308-1202
PART 2 - COPY TO BE RETURNED WITH RESPONSE

The dating stamp of the Patent Office on this card will be taken as an indication that the accompanying paper was filed.

Enclosed:

Sequence Amendment (21 pp.)
Sequence Listing (5 pp)
disk
Copy of Notice to File Missing Parts of Nonprovisional Application (2 pp.

Applicant(s) Jonathan W. JARVIK

Serial No. 09/788,269

Paper dated 06/29/01

Atty's File No. 2087-010261

Amount of Check $-------

Initials BEJ/rmc

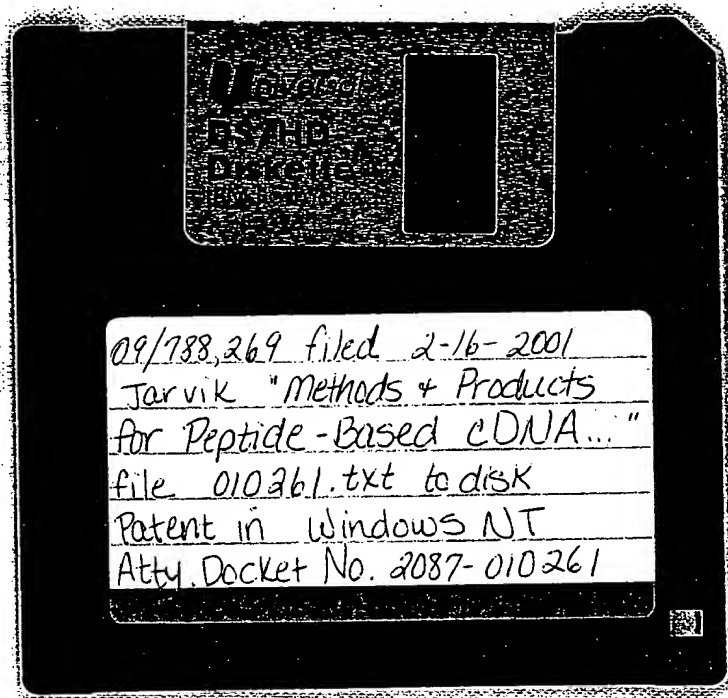OIPE JC102
MAR 1 8 2004
PATENT & TRADEMARK

OIPE JC42
JUL 0 3 2001
PATENT & TRADEMARK OFFICE
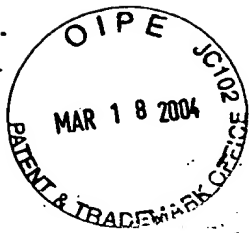
09/788,269 filed 2-16-2001
Jarvik "Methods & Products
for Peptide-Based cDNA..."
file 010261.txt to disk
Patent in Windows NT
Atty. Docket No. 2087-010261

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Group Art Unit 1645 :

In re application of : :

Jonathan W. JARVIK : **METHODS AND PRODUCTS FOR**
: **PEPTIDE-BASED cDNA**
Serial No. 09/788,269 : **CHARACTERIZATION AND ANALYSIS**

Filed February 16, 2001 :

Examiner Not Yet Assigned :

Pittsburgh, Pennsylvania
June 29, 2001

SEQUENCE AMENDMENT

Box SEQUENCE
Commissioner for Patents
Washington, DC 20231
Sir:
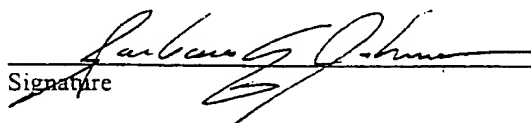
Pursuant to 37 C.F.R. § 1.821 et seq., Applicant submits the following Sequence

Listing and corresponding computer readable form (CRF) for insertion into the specification. A

copy of the Notice to File Missing Parts of Nonprovisional Application is also enclosed.

IN THE SPECIFICATION:

Please insert the attached Sequence Listing into the above-identified patent

application.

I hereby certify that this correspondence is being deposited with the
United States Postal Service as First Class Mail in an envelope
addressed to Commissioner for Patents, Washington, D.C. 20231 on
June 29, 2001.

Barbara E. Johnson, Registration No. 31,198
(Name of Registered Representative)

_____ 06/29/2001
Signature   Date

On pages 7-8, please delete paragraph 0018, and insert the following replacement paragraph 0018. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0018] If the nucleotide sequence is random, the probability that a sequence of given length translated from it will have a particular amino acid sequence can be calculated simply by multiplying together the frequencies in the genetic code of the codons encoding each amino acid in the sequence. Since some amino acids have as many as six codons and others as few as one, the predicted frequency will vary depending on the amino acid sequence itself. Thus the sequence LRRLLR (SEQ ID NO: 1), made up entirely of six-codon amino acids, will appear at a frequency of 1 in $(6/61)^6$, or approximately once in a million codons, and the sequence MWWMMW (SEQ ID NO: 2), made up entirely of one-codon amino acids, will appear at a frequency of 1 in $(1/61)^6$, or approximately once in fifty billion codons. The frequencies of other sequences will fall between these two extremes. The important point for us is that even a relatively short sequence will appear very rarely, and so if we can determine the amino acid sequence of a peptide translated from unknown sequence, we can match it to a portion of the reference sequence with high specificity.

On page 15, please delete paragraph 0030, and insert the following replacement paragraph 0030. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0030] Comparison of the experimental results with the values in the table indicates reveals a match to the predicted mass value for one of the ten candidates – specifically the sequence that begins at position 3190 of the reference sequence and proceeds from right to left. Retrieval of the reference sequence beginning at position 3190 indicates that the cloned sequence begins with "GAATTCTTACACCTCATACTTTCCCAAGCCCCAACTTTCTCATCT GAAAATGGTAATAGTATCATCCTTACATGTTTAAGGTCATGAATTGCTAT

GTGTA......(1st 100 nucleotides shown) (SEQ ID NO: 3). The identification is confirmed by dideoxy sequencing from a primer 150 nucleotides upstream of the junction between the pUC19 sequence and the EcoRI fragment.

On page 15, please delete paragraph 0032, and insert the following replacement paragraph 0032. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0032] The peptide TMITPSLHACRSTLED (SEQ ID NO: 4), representing the N-terminal 16 amino acids of the alpha-complementing factor of beta-galactosidase encoded in pUC19 (and also representing the 16 constant N-terminal amino acids in all of the peptides described in Example 1 above) is used to raise a polyclonal rabbit antibody using standard procedures.

On page 16, please delete paragraph 0034, and insert the following replacement paragraph 0034. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0034] The mass spectrum of the immunoprecipitate from the induced cell lysate of the clone under examination is observed to contain a distinct peak, at a position corresponding to a mass of 8485±3 Daltons, that is not observed in the control. Comparison of the experimental results with the values in the table in example 1 above indicates that the insert begins at position 9241 of the reference sequence and proceeds from left to right in the Genbank sequence. Retrieval of the reference sequence beginning at position 9241 indicates that the cloned sequence begins with GAATTCACATAAATCGCAAATTTTTTTTTCCTTCCCAGAGCC ATCCAAAACTCTGTTTGTCAAAGGCCTGTCTGAGGATACCACTGAAGAGA

CATTAAAG......(1st 100 nucleotides shown) (SEQ ID NO: 5). The identification is confirmed by dideoxy sequencing as described in Example 1.

On page 17, please delete paragraph 0037, and insert the following replacement paragraph 0037. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0037] To identify the nucleotide sequence adjacent to the pTriplEx' vector, each EcoRI site in the JO5584 sequence is identified and ligated, in silico, to the EcoRI site in the pTriplEx' vector. For each such in silico construct, the amino acid sequences of the two expected hybrid translation products (from each of the start codons in the vector to the first in frame stop codons encountered in the insert) are calculated. The mass of each peptide is calculated and all 10 peptide pairs are tabulated, as shown in the table below. Comparison of the experimental results (i.e., peptides of 4255 and 2635 Da.) with the values predicted in the table indicates that the insert begins at position 4028 of the reference sequence and proceeds in the forward direction. It is concluded that the 5' end of the sequence joined to the vector is GAATTCTCTTGGGTT TTGTGGTGTGCTAGACTTAATTACCCATGAATGATTT TGTCCTCTTGAGAAAATTTCAATAGCACATCTATTAGTGTTTTTTAT....(1st 100 nucleotides shown) (SEQ ID NO: 6). The identification is confirmed by dideoxy sequencing from the plasmid using a primer 150 nucleotides 3' to the pTriplEx' EcoRI site.

Position of EcoRI site	Orientation in pTriplEx'	Start Codon	Predicted Peptide Mass
3190	forward	1st	6137
3190	forward	2nd	5707
3190	reverse	1st	6278
3190	reverse	2nd	3891
4208	forward	1st	4255
4208	forward	2nd	2635
4208	reverse	1st	19748
4208	reverse	2nd	3905

-4-

6066	forward		1st	3595
6066	forward		2nd	3606
6066	reverse	1st	6401	
6066	reverse	2nd	1363	
9241	forward		1st	3583
9241	forward		2nd	7122
9241	reverse	1st	4582	
9241	reverse	2nd	1746	
9543	forward		1st	5306
9543	forward		2nd	1477
9543	reverse	1st	9906	
9543	reverse	2nd	2516	

The mass values above are computed by translating each hypothetical fusion polypeptide without the N-terminal methionine that is removed in vivo in E. coli.

On page 19, please delete the paragraph 0040, and insert the following replacement paragraph 0040. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0040] Two oligonucleotide primers are synthesized using standard methods. In one, CCCGAATTCAGCAGGTAAAAATCAAGG (SEQ ID NO: 7), the first 10 nucleotides contain an EcoRI site (underlined) and last 17 nucleotides correspond to the first 17 nucleotides of exon 2 of the human nucleolin gene. The other, GGGGAATTCTTACTCTTCTCCACTGCTAT (SEQ ID NO: 8), the last 17 nucleotides correspond to the reverse complement of the last 17 nucleotides of exon 2, followed immediately (in the sense orientation of the oligonucleotide) by the stop codon TAA and a sequence that includes an EcoRI site (underlined).

On pages 21-24, please delete paragraph 0044, and insert the following replacement paragraph 0044. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0044] The program was run with the 24 nucleotide input sequence CAACTAGAAGAGGTAAGAAACTAT (SEQ ID NO: 9). Two reading frames were selected; the forward reading frame beginning with the first nucleotide (F1) and the reverse (antisense) reading frame beginning with the second antisense nucleotide (R2). The results are shown below.

[begin]

Enter Sequence:

[input]   CAACTAGAAGAGGTAAGAAACTAT (SEQ ID NO: 9)

[output]   Protein: QLEEVRNY (SEQ ID NO: 10)

Which reading frames would you like to examine?

1: Forward (F1)

2: Forward; first base removed (F2)

3: Forward; second base removed (F2)

4: Reverse (R1)

5: Reverse first base removed (R2)

6: Reverse second removed (R3)

[input]   1,5

[output]   **MASS DIFFERENCES**

Location	Mutation	Frame F1	Frame R2
	None	1032.13	722.89
	/A(K)	0.04	0.00
1	C-{ G(E)	0.99	0.00
	\T(Z)	-1032.13	0.00
	/G(R)	28.06	0.00
2	(Q) A-{ T(L)	-14.97	0.00
	\C(P)	-31.01	0.00

	/G(Q)	0.00	0.00
3	A-{ T(H)	9.01	0.00
	\C(H)	9.01	0.00
	/A(I)	0.00	276.34
4	C-{ G(V)	-14.03	276.34
	\T(L)	0.00	0.00
	/C(P)	-16.04	299.37
5	(L) T-{ A(Q)	14.97	226.32
	\G(R)	43.03	200.24
	/G(L)	0.00	241.29
6	A-{ T(L)	0.00	241.33
	\C(L)	0.00	242.28
	/T(Z)	-790.84	-34.02
7	G-{ C(Q)	-0.99	-34.02
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-60.10
8	(E) A-{ T(V)	-29.99	16.00
	\C(A)	-58.04	-44.04
	/G(E)	0.00	-34.02
9	A-{ T(D)	-14.03	-34.02
	\C(D)	-14.03	-48.05
	/T(Z)	-661.72	0.00
10	G-{ C(Q)	-0.99	0.00
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-16.04
11	(E) A-{ T(V)	-29.99	23.98
	\C(A)	-58.04	43.03
	/T(D)	-14.03	0.00
12	G-{ C(D)	-14.03	-14.03

	\A(E)	0.00	34.02
	/T(L)	14.03	-423.52
13	G-{ C(L)	14.03	-423.52
	\A(I)	14.03	0.00
	/C(A)	-28.05	-60.04
14	(V) T-{ A(E)	29.99	-16.00
	\G(G)	-42.08	-76.10
	/G(V)	0.00	-26.04
15	A-{ T(V)	0.00	-49.08
	\C(V)	0.00	-48.09
	/G(G)	-99.14	0.00
16	A-{ T(Z)	-433.47	0.00
	\C(R)	0.00	0.00
	/T(I)	-43.03	76.10
17	(R) G-{ C(T)	-55.09	16.06
	\A(K)	-28.02	60.10
	/G(R)	0.00	10.04
18	A-{ T(S)	-69.11	14.02
	\C(S)	-69.11	-16.00
	/G(D)	0.99	0.00
19	A-{ T(Y)	49.08	0.00
	\C(H)	23.04	0.00
	/G(S)	-27.02	-28.05
20	(N) A-{ T(I)	-0.94	15.96
	\C(T)	-13.00	-42.08
	/A(K)	14.07	48.05
21	C-{ G(K)	14.07	14.03
	\T(N)	0.00	14.03

	/C(H)	-26.04	18.03	
	\G(D)	-49.08	0.00	
22	T-{ A(N)	-48.09	0.00	
	/G(C)	-60.04	-12.06	
23	(Y) A-{ T(F)	-16.00	15.01	
	\C(S)	-76.10	43.03	
	/C(Y)	0.00	-14.03	
24	T-{ A(Z)	-163.18	0.00	
	\G(Z)	-163.18	0.00	

Enter the detection threshold:

[input]   0.8 Dalton.

[output]   Undetectable amino acid substitutions: 1.(Q)C-A(K)

On page 26, please delete paragraph 0048, and insert the following replacement

paragraph 0048. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement

paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0048]   The sequence of exon 2 of the human rds/peripherin gene (Genbank
accession M73531) is shown below. Intron sequence is shown in lower case; exon sequence in
upper case.

gggaagcccatctccagctgtctgtttccctttaagTCGAATCAAGAGCAACGTGGATGGGCG

GTACCTGGTGGACGGCGTCCCTTTCAGCTGCTGCAATCCTAGCTCGCCACGGCCCTGC

ATCCAGTATCAGATCACCAACAACTCAGCACACTACAGTTACGACCACCAGACGGAG

GAGCTCAACCTGTGGGTGCGTGGCTGCAGGGCTGCCCTGCTGAGCTACTACAGCAGCC

TCATGAACTCCATGGGTGTCGTCACGCTCCTCATTTGGCTCTTCGAGgtaggccctgggcagctg

ggggtagagggtaaggagagcctcc (SEQ ID NO: 11)

-9-

On pages 26-27, please delete paragraph 0049, and insert the following replacement paragraph 0049. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0049] Two primers, of sequences GGCCCGGAATTCTCCAGCTGTCTGTTTCCCTTTAAG (SEQ ID NO: 12) and AATTTACTCGAGCTACCCCCAGCTGCCCAGGGCCTAC (SEQ ID NO: 13) were synthesized and used to PCR amplify rds/peripherin exon 2 from an individual known to carry a wild type allele of rds/peripherin. The amplicon was cut with EcoRI and XhoI and cloned into the EcoRI/XhoI sites of the pGEX derivative described in Nelson et al. The resulting plasmid was cut with Xho 1, treated with Klenow fragment of DNA polymerase, and self-ligated to produce a construct expected to produce a fusion protein with the sequence shown below.

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLE

FPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSK

DFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLD

AFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLIEGRGIQDLVPH

TTPHHTTPHHTTPHHTTPQDLNSPAVCFPLSRIKSNVDGRYLVDGVPFSCCNPSSPRPCIQY

QITNNSAHYSYDHQTEELNLWVRGCRAALLSYYSSLMNSMGVVTLLIWLFEVGPGQLGV

ARSSGRIVTD (SEQ ID NO: 14)

On page 27, please delete paragraph 0053, and insert the following replacement paragraph 0053. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0053] The amplicons described in the previous example are reamplified using the upstream primer

5'GGATCCTAATACGACTCACTATAGGGAGACCACC<u>ATG</u>CATCACCATCATCACCATCA

CCACTCTCCAGCTGTCTGTTTCCCTTTAAG (SEQ ID NO: 15) and the downstream primer

5' CTTAGTCATTATACCCCCAGCTGCCCAGGGCCTAC (SEQ ID NO: 16). The upstream

primer contains a T7 promoter followed by a translation initiation sequence (start codon

underlined) followed by a sequence encoding eight histidines followed by sequence identical to

the red/peripherin sequence immediately 5' to rds/peripherin exon 2.   The downstream primer

contains two stop codons (in antisense orientation) preceding the sequence complimentary to the

sequence just 3' to red/peripherin exon 2.

On page 30, please delete paragraph 0061, and insert the following replacement

paragraph 0061. Pursuant to 37 C.F.R. § 1.121, the following is a clean copy of the replacement

paragraph. A marked-up copy of the replacement paragraph is attached on separate sheets.

[0061]   Because the primers are all anchored by non-T nucleotides at their 3'

ends, only three of them will prime a given cDNA sequence. In the case of the hemoglobin alpha

2 transcript, which ends in the sequence GCGGCAAAAAAAAAAAAAAAAAAAAAAAA...,

(SEQ ID NO: 17) the primers that are extended are those ending in G.

REMARKS

Pursuant to the requirements of 37 C.F.R. §§ 1.821-1.825, Applicant submits the enclosed Sequence Listing and computer readable form (CRF). The amino acid sequences disclosed in the specification and drawings may be found in computer readable form in file 010261.txt on the enclosed diskette and are presented in the paper copy of the Sequence Listing, enclosed.
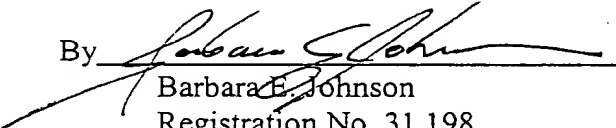
Applicant hereby certifies that the information recorded in computer readable form (CRF) supplied on the enclosed diskette as file 010261.txt is identical to the written Sequence Listing. The material presented in computer readable form is not new matter because it presents sequences the same as those disclosed in the specification, as filed.

Applicant believes that the requirements of 37 C.F.R. §§ 1.821-1.825 have been met.

Respectfully submitted,

WEBB ZIESENHEIM LOGSDON
ORKIN & HANSON, P.C.

By _____
Barbara E. Johnson
Registration No. 31,198
Attorney for Applicant
700 Koppers Building
436 Seventh Avenue
Pittsburgh, PA 15219-1818
Telephone: 412-471-8815
Facsimile: 412-471-4094

MARKED-UP AMENDED SPECIFICATION PARAGRAPHS

[0018]   If the nucleotide sequence is random, the probability that a sequence of given length translated from it will have a particular amino acid sequence can be calculated simply by multiplying together the frequencies in the genetic code of the codons encoding each amino acid [amino acid] in the sequence. Since some amino acids have as many as six codons and others as few as one, the predicted frequency will vary depending on the amino acid sequence itself. Thus the sequence LRRLLR (SEQ ID NO: 1), made up entirely of six-codon amino acids, will appear at a frequency of 1 in $(6/61)^6$, or approximately once in a million codons, and the sequence MWWMMW (SEQ ID NO: 2), made up entirely of one-codon amino acids, will appear at a frequency of 1 in $(1/61)^6$, or approximately once in fifty billion codons. The frequencies of other sequences will fall between these two extremes. The important point for us is that even a relatively short sequence will appear very rarely, and so if we can determine the amino acid sequence of a peptide translated from unknown sequence, we can match it to a portion of the reference sequence with high specificity.

[0030]   Comparison of the experimental results with the values in the table indicates reveals a match to the predicted mass value for one of the ten candidates -- specifically the sequence that begins at position 3190 of the reference sequence and proceeds from right to left. Retrieval of the reference sequence beginning at position 3190 indicates that the cloned sequence begins with "GAATTCTTACACCTCATACTTTCCCAAGCCCCAACTTTCTCATCT GAAAATGGTAATAGTATCATCCTTACATGTTTAAGGTCATGAATTGCTAT GTGTA......(1st 100 nucleotides shown) (SEQ ID NO: 3). The identification is confirmed by dideoxy sequencing from a primer 150 nucleotides upstream of the junction between the pUC19 sequence and the EcoRI fragment.

[0032]   The peptide TMITPSLHACRSTLED (SEQ ID NO: 4), representing the

N-terminal 16 amino acids of the alpha-complementing factor of beta-galactosidase encoded in

pUC19 (and also representing the 16 constant N-terminal amino acids in all of the peptides

described in Example 1 above) is used to raise a polyclonal rabbit antibody using standard

procedures.

[0034]   The mass spectrum of the immunoprecipitate from the induced cell

lysate of the clone under examination is observed to contain a distinct peak, at a position

corresponding to a mass of 8485±3 Daltons, that is not observed in the control. Comparison of

the experimental results with the values in the table in example 1 above indicates that the insert

begins at position 9241 of the reference sequence and proceeds from left to right in the Genbank

sequence.   Retrieval of the reference sequence beginning at position 9241 indicates that the

cloned   sequence   begins   with
GAATTCACATAAATCGCAAATTTTTTTTTCCTTCCCAGAGCC
ATCCAAAACTCTGTTTGTCAAAGGCCTGTCTGAGGATACCACTGAAGAGA

CATTAAAG......(1st 100 nucleotides shown) (SEQ ID NO: 5). The identification is confirmed by

dideoxy sequencing as described in Example 1.

[0037]   To identify the nucleotide sequence adjacent to the pTriplEx' vector, each

EcoRI site in the JO5584 sequence is identified and ligated, in silico, to the EcoRI site in the

pTriplEx' vector. For each such in silico construct, the amino acid sequences of the two expected

hybrid translation products (from each of the start codons in the vector to the first in frame stop

codons encountered in the insert) are calculated. The mass of each peptide is calculated and all

10 peptide pairs are tabulated, as shown in the table below. Comparison of the experimental

results (i.e., peptides of 4255 and 2635 Da.) with the values predicted in the table indicates that

the insert begins at position 4028 of the reference sequence and proceeds in the forward direction.

It is concluded that the 5' end of the sequence joined to the vector is
GAATTCTCTTGGGTT TTGTGGTGTGCTAGACTTAATTACCCATGAATGATTT
TGTCCTCTTGAGAAAATTTCAATAGCACATCTATTAGTGTTTTTAT....(1st 100

nucleotides shown) (SEQ ID NO: 6). The identification is confirmed by dideoxy sequencing from the plasmid using a primer 150 nucleotides 3' to the pTriplEx' EcoRI site.

Position of EcoRI site	Orientation in pTriplEx'	Start Codon		Predicted Peptide Mass
3190	forward		1st	6137
3190	forward		2nd	5707
3190	reverse	1st	6278	
3190	reverse	2nd	3891	
4208	forward		1st	4255
4208	forward		2nd	2635
4208	reverse	1st	19748	
4208	reverse	2nd	3905	
6066	forward		1st	3595
6066	forward		2nd	3606
6066	reverse	1st	6401	
6066	reverse	2nd	1363	
9241	forward		1st	3583
9241	forward		2nd	7122
9241	reverse	1st	4582	
9241	reverse	2nd	1746	
9543	forward		1st	5306
9543	forward		2nd	1477
9543	reverse	1st	9906	
9543	reverse	2nd	2516	

The mass values above are computed by translating each hypothetical fusion polypeptide without the N-terminal methionine that is removed in vivo in E. coli.

[0040] Two oligonucleotide primers are synthesized using standard methods. In one, CCCGAATTCAGCAGGTAAAAATCAAGG (SEQ ID NO: 7), the first 10 nucleotides contain an EcoRI site (underlined) and last 17 nucleotides correspond to the first 17 nucleotides of exon 2 of the human nucleolin gene. The other, GGGGAATTCTTACTCTTCTCCACTGCTAT (SEQ ID NO: 8), the last 17 nucleotides correspond to the reverse complement of the last 17 nucleotides of exon 2, followed immediately (in the sense orientation of the oligonucleotide) by the stop codon TAA and a sequence that includes an EcoRI site (underlined).

[0044] The program was run with the 24 nucleotide input sequence CAACTAGAAGAGGTAAGAAACTAT (SEQ ID NO: 9). Two reading frames were selected; the forward reading frame beginning with the first nucleotide (F1) and the reverse (antisense) reading frame beginning with the second antisense nucleotide (R2). The results are shown below.

[begin]

    Enter Sequence:

[input]   CAACTAGAAGAGGTAAGAAACTAT (SEQ ID NO: 9)

[output]   Protein: QLEEVRNY (SEQ ID NO: 10)

    Which reading frames would you like to examine?

    1: Forward (F1)

    2: Forward; first base removed (F2)

    3: Forward; second base removed (F2)

    4: Reverse (R1)

    5: Reverse first base removed (R2)

    6: Reverse second removed (R3)

[input]   1,5

[output]   MASS DIFFERENCES

Location	Mutation	Frame F1	Frame R2

	None	1032.13	722.89
	/A(K)	0.04	0.00
1	C-{ G(E)	0.99	0.00
	\T(Z)	-1032.13	0.00
	/G(R)	28.06	0.00
2	(Q) A-{ T(L)	-14.97	0.00
	\C(P)	-31.01	0.00
	/G(Q)	0.00	0.00
3	A-{ T(H)	9.01	0.00
	\C(H)	9.01	0.00
	/A(I)	0.00	276.34
4	C-{ G(V)	-14.03	276.34
	\T(L)	0.00	0.00
	/C(P)	-16.04	299.37
5	(L) T-{ A(Q)	14.97	226.32
	\G(R)	43.03	200.24
	/G(L)	0.00	241.29
6	A-{ T(L)	0.00	241.33
	\C(L)	0.00	242.28
	/T(Z)	-790.84	-34.02
7	G-{ C(Q)	-0.99	-34.02
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-60.10
8	(E) A-{ T(V)	-29.99	16.00
	\C(A)	-58.04	-44.04
	/G(E)	0.00	-34.02
9	A-{ T(D)	-14.03	-34.02
	\C(D)	-14.03	-48.05
	/T(Z)	-661.72	0.00

10	G-{ C(Q)	-0.99	0.00
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-16.04
11	(E) A-{ T(V)	-29.99	23.98
	\C(A)	-58.04	43.03
	/T(D)	-14.03	0.00
12	G-{ C(D)	-14.03	-14.03
	\A(E)	0.00	34.02
	/T(L)	14.03	-423.52
13	G-{ C(L)	14.03	-423.52
	\A(I)	14.03	0.00
	/C(A)	-28.05	-60.04
14	(V) T-{ A(E)	29.99	-16.00
	\G(G)	-42.08	-76.10
	/G(V)	0.00	-26.04
15	A-{ T(V)	0.00	-49.08
	\C(V)	0.00	-48.09
	/G(G)	-99.14	0.00
16	A-{ T(Z)	-433.47	0.00
	\C(R)	0.00	0.00
	/T(I)	-43.03	76.10
17	(R) G-{ C(T)	-55.09	16.06
	\A(K)	-28.02	60.10
	/G(R)	0.00	10.04
18	A-{ T(S)	-69.11	14.02
	\C(S)	-69.11	-16.00
	/G(D)	0.99	0.00
19	A-{ T(Y)	49.08	0.00
	\C(H)	23.04	0.00

	/G(S)	-27.02	-28.05
20	(N) A-{ T(I)	-0.94	15.96
	\C(T)	-13.00	-42.08
	/A(K)	14.07	48.05
21	C-{ G(K)	14.07	14.03
	\T(N)	0.00	14.03
	/C(H)	-26.04	18.03
	\G(D)	-49.08	0.00
22	T-{ A(N)	-48.09	0.00
	/G(C)	-60.04	-12.06
23	(Y) A-{ T(F)	-16.00	15.01
	\C(S)	-76.10	43.03
	/C(Y)	0.00	-14.03
24	T-{ A(Z)	-163.18	0.00
	\G(Z)	-163.18	0.00

Enter the detection threshold:

[input]   0.8 Dalton.

[output]   Undetectable amino acid substitutions: 1.(Q)C-A(K)

[0048]   The sequence of exon 2 of the human rds/peripherin gene (Genbank accession M73531) is shown below. Intron sequence is shown in lower case; exon sequence in upper case.

gggaagcccatctccagctgtctgtttccctttaagTCGAATCAAGAGCAACGTGGATGGGCG GTACCTGGTGGACGGCGTCCCTTTCAGCTGCTGCAATCCTAGCTCGCCACGGCCCTGC ATCCAGTATCAGATCACCAACAACTCAGCACACTACAGTTACGACCACCAGACGGAG GAGCTCAACCTGTGGGTGCGTGGCTGCAGGGCTGCCCTGCTGAGCTACTACAGCAGCC TCATGAACTCCATGGGTGTCGTCACGCTCCTCATTTGGCTCTTCGAGgtaggccctgggcagctg ggggtagagggtaaggagagcctcc (SEQ ID NO: 11)

[0049]   Two   primers,   of   sequences
GGCCCGGAATTCTCCAGCTGTCTGTTTCCCTTTAAG   (SEQ   ID   NO: 12)   and
AATTTACTCGAGCTACCCCCAGCTGCCCAGGGCCTAC (SEQ ID NO: 13) were synthesized and
used to PCR amplify rds/peripherin exon 2 from an individual known to carry a wild type allele of
rds/peripherin. The amplicon was cut with EcoRI and XhoI and cloned into the EcoRI/XhoI sites of the
pGEX derivative described in Nelson et al. The resulting plasmid was cut with Xho 1, treated with
Klenow fragment of DNA polymerase, and self-ligated to produce a construct expected to produce a
fusion protein with the sequence shown below.

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLE

FPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSK

DFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLD

AFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLIEGRGIQDLVPH

TTPHHTTPHHTTPHHTTPQDLNSPAVCFPLSRIKSNVDGRYLVDGVPFSCCNPSSPRPCIQY

QITNNSAHYSYDHQTEELNLWVRGCRAALLSYYSSLMNSMGVVTLLIWLFEVGPGQLGV

ARSSGRIVTD (SEQ ID NO: 14)

[0053]   The amplicons described in the previous example are reamplified using

the upstream primer

5'GGATCCTAATACGACTCACTATAGGGAGACCACCATGCATCACCATCATCACCATCA

CCACTCTCCAGCTGTCTGTTTCCCTTTAAG (SEQ ID NO: 15) and the downstream primer

5' CTTAGTCATTATACCCCCAGCTGCCCAGGGCCTAC (SEQ ID NO: 16). The upstream

primer contains a T7 promoter followed by a translation initiation sequence (start codon

underlined) followed by a sequence encoding eight histidines followed by sequence identical to

the red/peripherin sequence immediately 5' to rds/peripherin exon 2. The downstream primer

contains two stop codons (in antisense orientation) preceding the sequence complimentary to the sequence just 3' to red/peripherin exon 2.

[0061]   Because the primers are all anchored by non-T nucleotides at their 3' ends, only three of them will prime a given cDNA sequence. In the case of the hemoglobin alpha 2 transcript, which ends in the sequence GCGGCAAAAAAAAAAAAAAAAAAAAAAAA..., (SEQ ID NO: 17) the primers that are extended are those ending in G.
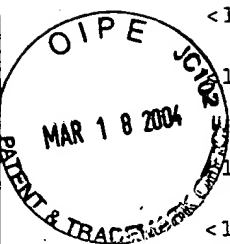
<110> Jarvik, Jonathan W.

<120> Methods and Products for Peptide-Based cDNA
      Characterization and Analysis

<130> 2087 010261

<140> US 09/788,269
<141> 2001-02-16

<150> US 60/182,983
<151> 2000-02-16

<160> 17

<170> Microsoft Word 97 SR-2

<210> 1
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Example of sequence made up entirely of six-codon amino acids

<400> 1
Leu Arg Arg Leu Leu Arg
 1   5

<210> 2
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Example of sequence made up entirely of one-codon amino acids

<400> 2
Met Trp Trp Met Met Trp
 1   5

<210> 3
<211> 100
<212> DNA
<213> Homo sapiens

<400> 3
gaattcttac acctcatact ttcccaagcc ccaactttct catctgaaaa tggtaatagt 60

atcatcctta catgtttaag gtcatgaatt gctatgtgta   100

<210> 4
<211> 16
<212> PRT
<213> Homo sapiens

<400> 4
Thr Met Ile Thr Pro Ser Leu His Ala Cys Arg Ser Thr Leu Glu Asp
 1   5   10   15

```
<210> 5
<211> 100
<212> DNA
<213> Homo sapiens

<400> 5
gaattcacat aaatcgcaaa ttttttttttc cttcccagag ccatccaaaa ctctgtttgt 60

caaaggcctg tctgaggata ccactgaaga gacattaaag                          100

<210> 6
<211> 99
<212> DNA
<213> Homo sapiens

<400> 6
gaattctctt gggttttgtg gtgtgctaga cttaattacc catgaatgat tttgtcctct 60

tgagaaaatt tcaatagcac atctattagt gttttttat                          99

<210> 7
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<221> SITE
<222> (4)..(9)
<223> Oligonucleotide primer containing EcoRI site

<400> 7
cccgaattca gcaggtaaaa atcaagg                                       27

<210> 8
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<221> SITE
<222> (4)..(9)
<223> Oligonucleotide primer containing EcoRI site

<400> 8
ggggaattct tactcttctc cactgctat                                     29

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Nucleotide input sequence used to deonstrate computer program
      capabilities

<400> 9
caactagaag aggtaagaaa ctat                                          24

<210> 10
<211> 8
```

<212> PRT
<213> Artificial Sequence

<220>
<223> Computer program output of encoded peptides

<400> 10
Gln Leu Glu Glu Val Arg Asn Tyr

<210> 11
<211> 326
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> (37)..(283)

<400> 11
gggaagccca tctccagctg tctgtttccc tttaagtcga atcaagagca acgtggatgg 60
gcggtacctg gtggacggcg tccctttcag ctgctgcaat cctagctcgc cacggccctg 120
catccagtat cagatcacca acaactcagc acactacagt tacgaccacc agacggagga 180
gctcaacctg tgggtgcgtg gctgcagggc tgccctgctg agctactaca gcagcctcat 240
gaactccatg ggtgtcgtca cgctcctcat ttggctcttc gaggtaggcc ctgggcagct 300
gggggtagag ggtaaggaga gcctcc 326

<210> 12
<211> 36
<212> DNA
<213> Artificial sequence

<220>
<223> Primer synthesized and used to PCR amplify rds/peripherin exon 2
      from an individual known to carry a wild type allele of
      rds/peripherin.

<400> 12
ggcccggaat tctccagctg tctgtttccc tttaag 36

<210> 13
<211> 37
<212> DNA
<213> Artificial sequence

<220>
<223> Primer synthesized and used to PCR amplify rds/peripherin exon 2
      from an individual known to carry a wild type allele of
      rds/peripherin.

<400> 13
aatttactcg agctaccccc agctgcccag ggcctac 37

<210> 14
<211> 364
<212> PRT
<213> Artificial sequence

<220>
<223> Fusion protein

<400> 14

```
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
1               5                   10                  15
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
        50                  55                  60
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
 65              70                  75                      80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
                85                  90                  95
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ile Glu Gly
        210                 215                 220
Arg Gly Ile Gln Asp Leu Val Pro His Thr Thr Pro His His Thr Thr
225                 230                 235                 240
Pro His His Thr Thr Pro His His Thr Thr Pro Gln Asp Leu Asn Ser
                245                 250                 255
Pro Ala Val Cys Phe Pro Leu Ser Arg Ile Lys Ser Asn Val Asp Gly
                260                 265                 270
Arg Tyr Leu Val Asp Gly Val Pro Phe Ser Cys Cys Asn Pro Ser Ser
        275                 280                 285
Pro Arg Pro Cys Ile Gln Tyr Gln Ile Thr Asn Asn Ser Ala His Tyr
        290                 295                 300
Ser Tyr Asp His Gln Thr Glu Glu Leu Asn Leu Trp Val Arg Gly Cys
305                 310                 315                 320
Arg Ala Ala Leu Leu Ser Tyr Tyr Ser Ser Leu Met Asn Ser Met Gly
                325                 330                 335
Val Val Thr Leu Leu Ile Trp Leu Phe Glu Val Gly Pro Gly Gln Leu
            340                 345                 350
Gly Val Ala Arg Ser Ser Gly Arg Ile Val Thr Asp
        355                 360
```

<210> 15
<211> 87
<212> DNA
<213> Artificial sequence

<220>
<221> misc_feature
<222> (35)..(37)
<223> Upstream primer used to reamplify amplicons
      Start codon at 35-37

<400> 15
ggatcctaat acgactcact atagggagac caccatgcat caccatcatc accatcacca 60
ctctccagct gtctgtttcc ctttaag   87

<210> 16
<211> 35
<212> DNA
<213> Artificial sequence

<220>
<223> Downstream primer used to reamplify amplicons

<400> 16
cttagtcatt ataccccag ctgcccaggg cctac   35

<210> 17
<211> 28
<212> DNA
<213> Artificial sequence

<220>
<223> Ending of hemoglobin alpha 2 transcript

<400> 17
gcggcaaaaa aaaaaaaaaa aaaaaaa   28